Factors affecting COVID-19 mortality: an exploratory study

Ashish Upadhyaya
School of Health Sciences, University of New Haven, West Haven, Connecticut, USA
Sushant Koirala
School of Medicine, Saint Louis University, St. Louis, Missouri, USA
Rand Ressler
College of Business Administration, Georgia Southern University, Statesboro, Georgia, USA, and
Kamal Upadhyaya
Department of Economics and Business Analytics, University of New Haven, West Haven, Connecticut, USA

Abstract
Purpose – The purpose of this paper is to study the factors affecting COVID-19 mortality.
Design/methodology/approach – An empirical model is developed in which the mortality rate per million is
the dependent variable, and life expectancy at birth, physician density, education, obesity, proportion
of population over the age of 65, urbanization (population density) and per capita income are explanatory
variables. Crosscountry data from 184 countries are used to estimate the quantile regression that is employed.
Findings – The estimated results suggest that obesity, the proportion of the population over the age of 65 and
urbanization have a positive and statistically significant effect on COVID-19 mortality. Not surprisingly, per
capita income has a negative and statistically significant effect on COVID-19 death rate.
Research limitations/implications – The study is based on the COVID-19 mortality data from June 2020,
which have constantly being changed. What data reveal today may be different after two or three months.
Despite this limitation, it is expected that this study will serve as the basis for future research in this area.
Practical implications – Since the findings suggest that obesity, population over the age of 65 and density
are the primary factors affecting COVID-19 death, the policy-makers should pay particular attention to these
factors.
Originality/value – To the authors’ knowledge, this is first attempt to estimate the factors affecting the
COVID-19 mortality rate. Its novelty also lies in the use of quantile regressions, which is more efficient in
estimating empirical models with heterogeneous data.
Keywords COVID-19, Mortality rate, Pandemic, Obesity, Urbanization
Paper type Research paper

Introduction
On December 31, 2019, Chinese officials in the city of Wuhan confirmed several cases of
pneumonia from an unknown cause [1]. As it quickly spread in Wuhan, scientists eventually
identified it as a novel coronavirus on January 7, 2020 [2]. According to the World Health
Organization (WHO), symptoms included dry cough, fever, shortness of breath, pneumonia

© Ashish Upadhyaya, Sushant Koirala, Rand Ressler and Kamal Upadhyaya. Published in Journal of Health Research. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licenses/by/4.0/legalcode

The authors did not receive financial support for the research, authorship or publication of this article.
and respiratory failure, which are symptoms similar to that of influenza and other viruses that attack the respiratory system. By January 11, 2020, the first virus-related death of a 61-year-old man in Wuhan was reported [1]. On January 20, the WHO reported the first confirmed cases of the novel coronavirus outside of China in Japan, South Korea and Thailand. As the number of infections and fatalities in China increased, on January 23rd, the Chinese Communist Party placed the entire population of Wuhan on mandatory lockdown [1]. All domestic travel in and out of the city was suspended except for international flights. On January 30th, the WHO declared a global public health emergency as a result of 9,000 cases reported in 18 countries, including China [1]. The following day, on 31st January, the US banned entry for most foreign nationals who had traveled to China [1].

The first coronavirus death outside China of a 44-year-old male, who was a resident of Wuhan city in China, was reported in the Philippines on February 2, 2020 [1]. By February 9th, the death toll in China rose to 811, surpassing the Severe acute respiratory syndrome (SARS) death in 2003. On February 11th, the WHO named the novel virus SARS-CoV-2 and its accompanying disease COVID-19 [2]. By the end of February, many other countries reported COVID-19 infections and fatalities, including South Korea, Iran, Italy and Japan. As a result, the US government banned flights from Iran, Italy, Japan and South Korea.

The first confirmed fatality related to COVID-19 in the United States was reported on February 29th, a man in his 50s in Washington state [1]. Over the next six weeks, infections and fatalities increased exponentially in Italy as well as many other European countries. By the end of March, infections, as well as fatalities, also started to rise exponentially in the United States. By June 2020, the death toll in the United States surpassed 100,000. Fatalities as a percentage of infection (i.e. case fatality rate) in European countries rose to alarming levels. The case fatality rate (CFR) in Asian and African countries, however, remained relatively low. In other words, the COVID-19 CFR and overall mortality rate, defined as the number of deaths in a population, varied from country to country.

Because COVID-19 is a new virus, there is a dearth of literature on its effect on the health outcome of the general population. To determine the socioeconomic determinants of the COVID-19 pandemic, a Bayesian model was developed and showed that the true empirical model of COVID-19 was constituted by only a few of its determinants [3]. There is other empirical literature [4, 5] that indicates that obesity, age over 65 and smoking are important factors causing COVID-19 death.

The purpose of this paper was to identify the factors correlated with the COVID-19 mortality rate. We hope to increase the level of understanding concerning factors that render some populations more vulnerable to the virus than others. As coronavirus continues its spread across the globe, even modest, but early advances in such knowledge could lead to a significant reduction in loss of life.

**Methodology**

Drawing upon the literature discussed above, the following model was developed in order to identify the factors affecting COVID-19 death.

\[
MRT = f(\text{PCGDP}, \text{DOC}, \text{OLD}, \text{OBE}, \text{URB}, \text{EDU}, \text{LEB})
\]  

(1)

where,

\[
\text{MRT} = \text{mortality rate per million of the population},
\]

\[
\text{OLD} = \text{proportion of population 65 years or older},
\]

\[
\text{OBE} = \text{percentage of the adult population with a BMI of 30 or greater},
\]

\[
\text{LEB} = \text{life expectancy at birth},
\]
DOC = availability of doctors per 1000 population,
URB = proportion of population living in urban areas,
PCGDP = per capita GDP in PPP (purchasing power parity) and
EDU = government expenditure on education out of total government spending.

In this estimation, the dependent variable MRT and the PCGDP controlled variable were transformed into a natural logarithm ln DRM and lnPCGDP, respectively. In Eqn (1), the coefficient OLD was expected to be positive as an increase in population over 65 years of age are more prone to die from pneumonia and respiratory failure caused by COVID-19. The coefficient of LEB could be positive or negative. Greater life expectancy at birth is an indication of good health of the population with a good immune system, which can fight the disease; therefore, LEB should have a positive coefficient. However, at the same time, greater life expectancy at birth implies people live longer with aging, and thus more prone to die of COVID-19 if infected. In this case, it can have a negative coefficient. DOC is expected to have a negative effect on the death rate as an increase in the number of medical doctors implied better healthcare for the patients infected and prone to COVID-19 infection and death.

Eqn (1) also included some social as well as economic variables. One of them was urbanization (URB). People who live in rural areas are more likely to maintain some sort of social distance. In contrast, those in urban areas, particularly large cities, often cannot maintain much, if any, social distance. This is particularly true for those who rely on public transportation. The increased population density present in urban areas almost certainly increased exposure to all communicable pathogens. As a result, urban residents were more likely to be infected with a heavy viral load which could increase the severity of COVID-19 and lead to death. The coefficient of this variable, therefore, was expected to be positive. The next variable in the model was per capita income (PCGDP). Higher per capita income implies better health and a higher quality of life. Therefore, we expected this variable to have a negative coefficient. Finally, a lack of education was also expected to harm the COVID-19 mortality rates, as a better-educated population was more informed about the prevention and treatment of COVID-19. However, a greater number of higher education institutions lead to an increased risk of transmission through contact with students, faculty and staff in communities where higher education institutions are located. This risk can increase the COVID-19 mortality rate; therefore, it is possible that EDU can have a positive coefficient.

A sample size of observations from 184 countries was used for this study. The data for mortality rate caused by COVID-19 was from June 22, 2020, or before and derived from Johns Hopkins University and Medicine Coronavirus Resource Center. The remaining data used in the model were derived from the CIA country fact book from 2020.

Ethical considerations
This study did not involve human beings during the research. The University of New Haven Institutional Research Board (IRB) exempted it under 45 CFR 46.104(b) (4) and approved it for dissemination (Protocol no. 2020-074).

Results
This investigation employed two different estimation techniques: ordinary least square (OLS) and quantile regression to estimate the death rate per million, which had been transformed to the natural logarithm while estimating with results presented in Table 1. The results from OLS are presented simply for reporting purposes, as the study discusses more on the results from the quantile regression.
<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS</th>
<th>Q.10</th>
<th>Q.25</th>
<th>Q.50</th>
<th>Q.75</th>
<th>Q.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT</td>
<td>-0.0850081</td>
<td>0.00000635</td>
<td>0.000019</td>
<td>0.0000178</td>
<td>-0.0000812</td>
<td>0.0000223</td>
</tr>
<tr>
<td></td>
<td>(0.4274728)</td>
<td>(0.00000739)</td>
<td>(0.0000153)</td>
<td>(0.0000237)</td>
<td>(0.00000848)</td>
<td>(0.0001421)</td>
</tr>
<tr>
<td>lnPCGDP</td>
<td>-0.0346124</td>
<td>3.27E-08</td>
<td>-0.000000386</td>
<td>-0.00000276*</td>
<td>-0.00000361</td>
<td>-0.00000692</td>
</tr>
<tr>
<td></td>
<td>(0.0228639)</td>
<td>(0.000000385)</td>
<td>(0.000000881)</td>
<td>(-0.00000152)</td>
<td>(0.000000645)</td>
<td>(0.00000974)</td>
</tr>
<tr>
<td>DOC</td>
<td>0.0287354</td>
<td>-1.48E-08</td>
<td>0.00000354</td>
<td>-0.0000136</td>
<td>0.00000209</td>
<td>-0.0000045</td>
</tr>
<tr>
<td></td>
<td>(0.0192335)</td>
<td>(0.000000611)</td>
<td>(0.000000849)</td>
<td>(-0.000000132)</td>
<td>(0.000000774)</td>
<td>(0.00000166)</td>
</tr>
<tr>
<td>OLD</td>
<td>-0.0081971</td>
<td>0.00000636</td>
<td>0.00000855</td>
<td>0.00000163***</td>
<td>0.00000613</td>
<td>0.0000248***</td>
</tr>
<tr>
<td></td>
<td>(0.0059247)</td>
<td>(0.000000611)</td>
<td>(0.000000849)</td>
<td>(0.000000132)</td>
<td>(0.000000774)</td>
<td>(0.00000166)</td>
</tr>
<tr>
<td>OBE</td>
<td>-0.0002805</td>
<td>0.000000031</td>
<td>0.00000125</td>
<td>0.00000089**</td>
<td>0.00000205</td>
<td>0.00000171</td>
</tr>
<tr>
<td></td>
<td>(0.00041219)</td>
<td>(0.0000000794)</td>
<td>(0.000000221)</td>
<td>(0.000000411)</td>
<td>(0.000000496)</td>
<td>(0.00000238)</td>
</tr>
<tr>
<td>URB</td>
<td>0.0008671</td>
<td>0.000000547**</td>
<td>6.29E-08</td>
<td>0.000000325**</td>
<td>0.00000129***</td>
<td>0.00000175**</td>
</tr>
<tr>
<td></td>
<td>(0.0016395)</td>
<td>(0.000000239)</td>
<td>(0.0000000864)</td>
<td>(0.000000414)</td>
<td>(0.000000443)</td>
<td>(0.000000889)</td>
</tr>
<tr>
<td>EDU</td>
<td>0.0210648</td>
<td>-6.92E-08</td>
<td>-9.27E-08</td>
<td>-1.54E-08</td>
<td>0.00000121</td>
<td>0.00000119*</td>
</tr>
<tr>
<td></td>
<td>(0.0123682)</td>
<td>(0.000000165)</td>
<td>(0.000000444)</td>
<td>(0.000000868)</td>
<td>(0.000000278)</td>
<td>(0.00000631)</td>
</tr>
<tr>
<td>LEB</td>
<td>0.0049888</td>
<td>0.000000063</td>
<td>0.00000254</td>
<td>0.00000163</td>
<td>0.00000813</td>
<td>-0.00000183</td>
</tr>
<tr>
<td></td>
<td>(0.00060193)</td>
<td>(0.000000102)</td>
<td>(0.000000199)</td>
<td>(0.000000308)</td>
<td>(0.00000121)</td>
<td>(0.00000275)</td>
</tr>
</tbody>
</table>

**Note(s):** N = 184
a) Standard errors are in the parentheses. b) Quantile regression standard errors are bootstrapping standard errors. c) *, ** and *** represent variables are significant at 10%, 5% and 1%, respectively.
The benefits of using quantile regression were that the model would not be restricted by any specific distribution of the residuals. Unlike OLS, which is a mean regression, quantile regression used the median and therefore determined the relationships between variables that were outside of the mean of the data. The quantile regression informed us about how the effect of changing X by 1 unit varied across the conditional distribution of Y. Hence, the use of quantile regression was useful in understanding outcomes that were nonnormally distributed and that had nonlinear relationships with predictor variables. Therefore, using a quantile regression for this study allowed us to capture every possible effect of variables on the dependent variable lnDRM. The Stata command `sqreg` performed simultaneous quantile regression, and without any option, the `sqreg` did a median regression in which the coefficients were estimated by minimizing the absolute deviations from the median [6, 7]. As it was an estimate of central tendency, the median regression was a resistant measure that was not as significantly affected by outliers as a mean regression would be. The standard error estimates provided by `sqreg` were computed with the bootstrap standard errors via bootstrapping rather than through analytical formula. Since the analytical technique treated the sample means as fixed, whereas the bootstrap estimates accounted for their sampling variability and the standard errors estimated via bootstrapping, it was better when compared to the standard that OLS provides [8].

The quantile regression was run for 10, 25, 50, 75 and 90% quantiles. Due to the significance seen at 50% quantile, the results primarily discussed were based on this quantile while the other quantiles were presented for comparison purposes only. The results from the quantile regression showed that the natural log of per capita gross domestic (lnPCGDP) was significant at 10%, suggesting that if the 50% quantile of the DRM was considered, a 1% increase in lnPCGDP would decrease the mortality rate per million due to COVID-19 by 2.76*10^{-06}%. It provided rationality of increasing the healthcare budget to reduce the death rate per million population from such pandemics. Another important finding of this study was that it quantified the effects of several human health-related factors. This study found that if a person was old (OLD), it increased the mortality rate due to COVID-19 by 1.63*10^{-04}% compared to the young with COVID-19 (p < 0.01). The predicted confidence interval of the effect of variable OLD increased gradually from 10 to 50% quantiles; the confidence interval gradually fanned out after reaching the 75% quantiles and above, which can be seen in Figure 1. Similarly, another health-related variable, obesity (OBE) was significant in contributing to an increase in the mortality rate per million in this study. The result shows that obesity has led to an increase in the mortality rate per million by 8.09*10^{-04}%.

Besides the human health-related variables, this study also tested the effect of urbanization (URB), which is a human-related activity in a developing society. This study found that the URB variable was significant at 5% with a positive sign suggesting that the effect of living in an urban environment resulted in an increase of mortality rate per million by 3.25*10^{-05}%. This URB variable also behaves with a similar pattern to OLD variables when the quantiles of the dependent variable MRT moves from 10 to 50% quantiles, and beyond, its predicted confidence level fanned out after reaching the 75th quantile.

In order to check the robustness of our study, we also estimated the model by replacing the dependent variable mortality rate per million with the CFR (death as a percentage of infected). The estimated result was consistent with our results in Table 1 with the exception that the variable OBE was not statistically significant in explaining the CFR. As in the case of our
Note(s): Quantile plots with a confidence interval of predicted variable mortality rate per million (MRM) and corresponding controlled variables. Predicted values are in red with a corresponding confidence interval at 95%
previous estimation, the proportion of the population over 65 and urbanization were statistically significant in explaining the CFR.

Discussion
The estimated results above suggest that obesity (OBE) is a significant variable explaining the COVID-19 mortality rate. Over the past four decades, obesity has developed into its own global epidemic [9]. In 2016, there were nearly two billion adults over the age of 18 who were clinically overweight, defined as a BMI $\geq 25$, an early three-fold increase since 1975 [10]. The statistical significance of the OBE variable in COVID-19 mortalities corroborates previous studies identifying clinical obesity, defined as a BMI $\geq 30$ in adults, as a critical factor in COVID-19 mortality rates [10–12]. Of the three most common clinical comorbidities identified in patients hospitalized with COVID-19 – obesity, hypertension and diabetes [13] – obesity was the most preventable at the individual and population levels. Obesity is linked to higher mortality rates in numerous diseases, including influenza viruses, which pose a significant threat to become a future epidemic or pandemic. A large-scale effort to reduce global obesity rates is critical for minimizing mortality rates during future pandemics. Pragmatic policies designed to encourage healthy lifestyle practices should be adopted by nations with high obesity rates.

A common government intervention method used at local and national levels is the use of public health campaigns. These campaigns should avoid emphasizing an ideal body size or shape, thereby unintentionally stigmatizing overweight and obese populations. Instead, they should advocate for health and wellness by promoting nutritious food options and encouraging regular exercise. Oftentimes, well-intended public health campaigns unknowingly feature stigmatizing images of overweight or obese people. Analysis of stigmatizing campaigns compared to neutral campaigns suggest that stigmatization does not lead to an increase in motivation [14]. Research suggests that public health campaigns, a common strategy to reduce and prevent obesity, are not an effective means of motivating long-term lifestyle changes [9].

Contrary to the long-held belief that high-fat foods are the primary driver of obesity, recent studies have linked excess sugar consumption to rising obesity rates. Sugar is a significant driver of obesity in the United States, which has the 11th highest obesity rate out of 196 countries analyzed by the WHO [10]. The United States consumes more than 300% of the recommended amount of added sugar on a daily basis [15]. Sugar can be listed as an ingredient using a variety of names such as, but not limited to dextrose, glucose, maltose, xylose or evaporated cane juice. The various names for sugar used by food manufacturers create an added challenge in the effort to reduce obesity rates. Furthermore, a 2017 meta-analysis on the effects of mandatory calorie disclosure to consumers dining at restaurants found that consumers across all demographics consumed an average of 27 fewer calories per meal. However, calorie disclosure on menus had a greater effect on obese consumers. Obese consumers chose meal options with an average of 67 fewer calories [16]. Policy changes at the regulatory level that increase consumer transparency may be necessary and a more effective means of obesity prevention. For example, labels on food and drink items containing added sugar could be mandated to state a message such as “This product contains added sugar, which may be linked to obesity and other health problems” [17]. Similar labels on cigarette packages are effective in educating smokers on the negative health effects of smoking [18].

Our empirical findings also suggest that senior citizens are more prone to COVID-19 death compared to the young population. In the European Union, the percentage of the population over 65 is over 20% [19]. In the United States, citizens over 65 comprise 16.2% of the total population [20], with some states such as Florida, Maine, Vermont and West
Virginia at or above 20%. In Japan, citizens over 65 comprise 30% of the population, expected to further increase in the future. The populations of Asia, Africa and Latin America are relatively young compared to Europe, Japan and other industrialized countries. This age discrepancy may be one reason why the COVID-19 mortality rate is lower in Asian and African countries relative to Europe and North America. Considering that an older population is more vulnerable to death from COVID-19 and other similar respiratory diseases such as influenza, the policy implication we can derive is that the central, state and local governments in every nation should prioritize the safety of this age group, particularly during epidemics and pandemics. Interestingly, Japan has the largest percentage of elderly citizens among its population. Despite that, it has one of the lowest COVID-19 mortality rates. The percentage of COVID-19 deaths occurring in Japanese nursing homes is 14% compared to 40% in the United States. In the long-term care center, the density of beds per 1000 elderly people is 53 in the United States and 12 in Japan [21]. Greater density facilitates the transmission of diseases. Also, Japanese culture, which emphasizes respect for elders, prioritizes minimizing or eliminating other infectious diseases from nursing homes and other long-term care facilities and also consequently minimized COVID-19 deaths. Japan serves as a role model to the rest of the world on how to protect the elderly.

Another variable that was statistically significant to COVID-19 mortality rates is a nation’s degree of urbanization, which captures population density pockets. The recent outbreak of COVID-19 in India, Brazil and South Africa demonstrate that the majority of COVID-19 deaths have been in dense urban areas such as New Delhi, Mumbai, Sao Paolo and Johannesburg. Similar results were seen in the United States; high-density pockets of New York City, Miami and New Orleans have a high COVID-19 mortality rate. If we analyze these fatalities carefully, we see that the living conditions of people in these areas include three or more people living in a small space, often a single room with poor sanitation conditions. These living conditions are conducive to community spread and high transmission rates. The situation is more critical in the favelas and slums of Sao Paolo and Mumbai. In the Mumbai slum of Dharavi, 67% of households rely on community toilets, while soap and clean water are scarce, and social distancing is impossible. Consequently, 40–60% of people living in Mumbai’s slums were infected in mid-July 2020 [22]. However, some cities such as Trivandrum, the capital city of the Indian state of Kerala, acted quickly to increase testing and contact tracing which resulted in one of the lowest infection and mortality rates in India, despite thousands of people returning from international locations [22]. Hong Kong, Tokyo and Ho Chi Minh City were also able to manage COVID-19 through lessons learned from the 2003 SARS experience. Widespread testing and contact tracing, quarantines and, if necessary, lockdowns were quickly implemented, successfully minimizing COVID-19 transmission rates. Similar policies that emphasize the quick implementation of evidence-based infectious disease mitigation tactics should be instituted in nations around the world.

Our estimated regression result suggests that the per capita income is negatively related to COVID-19 mortality rates. Higher-income countries have greater access to modern health-care systems whereas the poorest countries lack such facilities. The few facilities that do exist are typically reserved for the wealthiest citizens. A healthy population, regardless of socioeconomic standing, is necessary for a healthy nation. Governments must pass legislation that prioritizes building strong health-care infrastructure and maximize funding for multiple aspects of public health and disease prevention. Legislative items include a greater number of hospitals and primary care clinics, modern diagnostic and therapeutic technology, strong emphasis on healthcare and medicine during secondary education, equitable access to healthcare, health and nutrition education, better sanitation, a national stockpile of personal protective equipments (PPE) and technology for efficient contact tracing.
Conclusion
This paper estimates and analyzes the factors affecting the COVID-19 mortality rate and suggests policy recommendations based on the results. A model, using cross-sectional data from 184 countries, was developed in which the per capita income, education, availability of doctors per thousand population, life expectancy at birth, urbanization, the proportion of the population over the age of 65 and obesity were the dependent variables explaining the COVID-19 death rate per million. Since the data were not balanced, quantile regression was used, and the results derived from this estimation were used to infer the findings for analysis. The estimated results suggest that obesity, the proportion of the population over the age of 65 and urbanization as a measure of density had a statistically significant positive effect on the COVID-19 mortality rate. Per capita income, however, had a negative and statistically significant effect on the COVID-19 mortality rate.

There are considerable limitations to any study employing data aggregated to the country level. Ours is no exception. Such data will likely not capture factors relevant to our research question that more refined data would reveal. We tried to mitigate this weakness to some degree by using the quantile estimation approach. Still, estimates borne from cross-sectional, country-wide data should be considered with caution. In the case of the coronavirus pandemic, caution is also called for due to the continuing spread across the globe. The data revealed today may be different in two months. We recognize these shortcomings in our analysis. Despite these limitations, we hope that this study will serve as the basis for future research in this area.

References


Corresponding author
Kamal Upadhyaya can be contacted at: Kupadhyaya@newhaven.edu

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com